DOCUMENT RESUME

ED 202 583 PS 012 185

AUTHOR Walker, Arlene
TITLE Intermodal Perception of Affect Expressions by Infants.
PUB DATE Apr 81
NOTE 20p.; Paper presented at the Annual Meeting of the Midwestern Psychological Association (53rd, Detroit, MI, April 30 - May 2, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Affective Behavior; Attention; *Auditory Perception; Early Childhood Education; Films; *Infant Behavior; Infants; *Visual Perception
IDENTIFIERS Facial Configuration; *Intermodal Perception; Voice Qualities

ABSTRACT

Four experiments (E1, E2, E3 and E4) investigated whether or not 5- to 7-month-old infants could detect auditory-visual relationships in audiovisual presentations of affective expressions, thereby perceiving the bimodally-presented expressions as unitary events. In E1, 16 infants were simultaneously shown two 2-minute films of a "happy" and a "neutral-to-serious" woman. During the presentation of the two filmed facial expressions, a single sound track was played. Each infant viewed the two films twice, one time with the sound track from the happy film and once with the neutral sound track. Visual fixation was recorded. A second group of infants was shown the two filmed facial expressions simultaneously but without an accompanying sound track. Infants viewing the silent films showed no significant differences in attention to either of the films. Infants viewing the sound films showed a preference for the happy film but also looked longer at the film accompanied by sound, whether the image projected was happy or neutral. Variations of E1 were conducted in E2, E3 and E4: E2 eliminated the silent film condition and presented happy and sad expressions; E3 eliminated synchrony between lip movement and audible speech; and E4 presented happy and angry expressions, half of which were presented upside down. In general, results were as in E1, except that infants increased their looking time while viewing the asynchronous sound film and showed no difference in looking time when films were presented upside down. It is concluded that 5- to 7-month-old infants detect information that is invariant across optic and acoustic displays of a single, affective expression. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Arlene S. Walker
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED202583

Intermodal Perception of Affect Expressions by Infants

Paper presented at the Midwestern Psychological Association Meetings, 1981

Arlene Walker
Department of Psychology
Cornell University

PS012185

I have chosen to talk today about the perception of affective expressions by infants, specifically the perception of expressions as multimodal events. I am doing so for a number of reasons. First, it seems that most of the research done with very young human infants that pertains to expressive behaviors has concerned the development of affect in infants, and the facial expressions made by these infants. Research has focused on the infants' behavior and his or her influence on others, rather than on the infants' abilities to recognize and respond to affective expressions. Second, I'm emphasizing infants' inter modalperception of expressions because I think that by exploring their intermodal perception we can make some progress towards determining whether infants perceive expressions as meaningful behaviors, as well as explore the sophistication of infants' abilities to detect intermodal information. The research that has investigated infants' perceptions of expressions has asked primarily whether infants can discriminate between various facial expressions (e.g., Wilcox and Clayton, 1968; Young-Browne, Rosenfeld and Horowitz, 1977). The results have indicated that infants can discriminate a few facial expressions as depicted in photographs or slides, but they tell us little about the developing perception of expressive behaviors. Evidence that infants can differentiate facial, vocal, or gestural expressions does not tell us whether these behaviors are perceived as meaningful--either as

information about another person's internal state, or as information about another person's likely behavior [the "affordance" of an expression (Gibson, 1979)].

A young infant typically observes unified expressive behaviors; that is he usually sees a person's face and gestures, hears that person's voice and (often) is touched by that person. Rarely is an infant presented with only a voice or only a face and gestures when someone is interacting with him. Infants do hear voices emanating from adjoining rooms, of course, but these vocalizations are usually directed to someone else. It is quite possible that a very young infants needs all of the information available multimodally for accurate discrimination and comprehension to occur. In these expressive events there is invariant information given optically and acoustically. Adults clearly detect this information and perceive a unified, meaningful expression. If infants detect this invariant information, they may be perceiving something abstract and amodal in that expressive behavior as well. Once it can be demonstrated that infants detect intermodal invariants, we can make further attempts to determine whether these are responded to in appropriate and differential ways.

The experiments I am going to describe now investigated young infant's perceptions of bimodally presented expressions. As already previewed, they asked whether infants could detect information that was invariant across the optic and acoustic presentations of a single affective expression. I used a method developed by Elizebeth Spelke (1976) for each of the four experiments I will be discussing. Spelke showed that when infants are presented with two filmed events side by side while one of the corresponding soundtracks is played, the infants look primarily at the sound-matched film, evidently detecting the correspondence. Thus far this method has been used to examine

infants' detection of invariant temporal structure or common spatial location underlying the correspondence. The present experiments attempted to investigate infants' detection of more abstract, higher-order relations found across the optic and acoustic presentations of the same event, specifically the same affective expression. The first 2 experiments merely demonstrate that infants can detect intermodal relations in optic and acoustic presentations of expressions; the latter 2 begin to explore what those intermodal relationships might be.

Experiment 1

In the first experiment, then, sixteen seven-month-old infants were shown two films, rear-projected side by side on a screen for two minutes. One film depicted a woman (from the shoulders up) smiling, laughing, and talking continuously, the "happy" film. The other film showed the same woman speaking with a serious expression on her face, also speaking continuously, the "neutral-to-serious" film. The soundtracks were recorded on the film when the movies were made, so they were in synchrony with the woman's face and lip movements. (This becomes an important point, for synchrony relations are detected by very young infants). The infant's attention was attracted to the center of a split screen, and the projectors were turned on. During the presentation of the two filmed facial expressions, a single soundtrack, appropriate to one of the facial expressions, was played through a centrally-placed speaker.

Each infant viewed the two films twice, one time with the soundtrack from the happy film and once with the neutral soundtrack. Visual fixation was recorded. I hypothesized that if the infants could detect the relationship between the sight and sound of an expression, they would look preferentially at a facial expression when it was accompanied by its soundtrack.

A second group of infants were shown the 2 filmed facial expressions side by side but without a soundtrack accompanying them.

The infants who viewed the films silently showed no significant differences in looking time to either of these films. Proportions of fixation to the happy film and to the right side films were .551 [t (11) .953, p <.361 and .565 [t (11) = 12.4, p <.240], respectively.

In contrast, the infants in the single soundtrack condition looked differentially to the filmed expressions. They markedly increased their looking time to a film when it was in sound, compared to when it was silently projected alongside the other, sound-specified film, as shown in Figure 1.

The average proportion of TLT directed at the neutral film was greater when it was presented with its soundtrack, .371 compared to .457 [t (15) = 2.18, p <.046]. Similarly, the proportion of TLT to the happy film was greater when it was accompanied by its soundtrack [.543 compared to .629, t (15) = 2.24, p <.04]. In general, the infants looked longer at the happy film (whether it was silent or sound-specified): that proportion averaged .585 [t (31) = 3.53, p <.001]. TLT to the 2 films averaged 51 seconds (SD = 21.9) in a session.

The infants responded to the sound-specified films by visually attending to them in this experiment. They also showed a preference for the happy facial expression in the single soundtrack condition. It is not the case, however, that the infants merely looked more to the happy face when any soundtrack co-occurred, for the infants increased their proportions of looking time to a film when it was in sound, as compared to when it was silent, for both happy and neutral films. But the infants did not show preferential looking in the no soundtracks condition. Although the absence of a preference does

not preclude discrimination (especially with infant subjects) it may be that the voice accompaniment rendered the filmed expressions more discriminable or the sound-specified film more salient in the single soundtrack condition. The results suggest that the infants detected some information that was invariant across the visible facial expression and its vocal counterpart and looked accordingly.

Experiment 2

These results, however, may be limited to (1) the particular expressions selected for filming, (2) the films themselves, and/or (3) to the particular age group participating. A second experiment, then, was conducted, aimed at testing the generality of these prior resluts, and demonstrating similar perceptual abilities in younger infants. In this experiment, five-month-old and seven-month-old infants participated. All the procedures were the same, except that there was no silent-only group. These infants were shown new films--another happy, and a sad filmed expression--side by side along with a single soundtrack. The results were the same as before. Both the five- and seven-month-old infants increased their looking time to a film when it was a sound, compared to when it was silent. They also had an overall preference for the happy film whether it was silent or sound-specified. Again the pattern of looking suggests that the infants perceived information that was invariant across the optic and acoustic presentations of an affective expression (see Figure 1). They responded by looking proportionately longer at that film, despite an overall preference for the filmed happy expression.

Experiment 3

The infants in these 2 experiments clearly extracted visual and auditory information about a person (or a film of a person), and they were able to

determine which of two visible persons was speaking. But while these first experiments indicated that infants of five and seven months detect the correspondence between the optic and acoustic presentations of a single affective expression, they do not tell us to what invariant information the infants were responding. Young infants are very sensitive to intermodal relations. Spelke and her colleagues, and Dodd in England have demonstrated that infants as young as four months can detect the synchrony between lip movements and heard speech. Much earlier Allen, Walker, Symonds, and Marcell found that infants of about seven months will find a simple rhythm to be equivalent across modalities. And recently I demonstrated, in collaboration with Lorraine Bahrick and Dick Neisser, that infants of four months can use intermodal relationships to visually follow one event, even when it is optically superimposed on other ongoing action. In that series of studies infants clearly followed one of the events to the exclusion of the other, apparently on the basis of amodal, invariant relationships which served to separate the sound-appropriate film from the other event (that did not share this structure). Temporal synchrony, common temporal pattern and rate of action, as well as appropriateness of the audible sounds to the visible motions all could have played a role in the infants' selectivity.

In the 2 experiments discussed so far on infants' perceptions of expressions, several kinds of information are invariant across the displays as well. Most obvious are temporal synchrony relations. The lip movements and vocal structure are synchronous, and the facial and vocal expressions share a common, if irregular, tempo. In addition, what might be called "intensity" relations may be evident in the displays. For example, the animation visible in the

happy facial expression is paralleled by the wide range of pitch and loudness that occurs in the happy vocal expression. And finally, the displays have a common affective tone since, of course, they are merely components of the same expression. The infants may have responded to any or all of this information. For this reason, I conducted a third experiment in which synchrony between the lip movements and audible speech was eliminated to determine whether synchrony alone could account for the infants' preferential looking.

Sixteen seven-month-old infants were shown the happy and neutral films that were used in the first experiment. The soundtrack from each film was tape-recorded and played through a centrally-placed speaker. The speech was delayed by approximately five seconds in relation to the ongoing, appropriate facial expression, making it very unlikely that the infants could resolve the asynchrony. All other procedures were identical to those used before.

In this experiment the discrepancy produced by the asynchronous vocal expression resulted in greater fixation overall. These infants spent a greater number of seconds looking at the pair of films than did the infants in Experiment 1 who were of the same age and were presented with the same films [76 seconds compared to 51 seconds, $t\ (54) = 3.21$, $p < .003$]. In general, the infants increased their looking time to a film when it was sound-specified (although asynchronous) compared to when it was not [neutral, .470 compared to .381, $t\ (15) = 1.93$, $p = .035$, one tailed; happy, .620 compared to .530, $t\ (15) = 1.94$, $p = .035$]. As in the previous experiments, the infants showed an overall preference for the happy film as well [.575, $t\ (31) = 2.66$, $p < .02$].

A closer look at these data disclosed that the preferences for the sound film only emerged in the second session, however. You will remember

that the infants saw the pair of films twice, once with one soundtrack for 2 minutes and then with the other soundtrack. During the first session the infants spent only 49% of the time looking to the sound-matched, asynchronous happy and neutral films. During the second session, however, this averaged 60%, a significant increase [t (31) = 1.82, p <.04, one tailed]. Preferences for the sound-matched film appeared during the second 2 minute session and increased steadily across that session, such that the infants were looking to the sound-matched happy and neutral films during the last 30 seconds, 76% (for happy) and 63% (for neutral) of the time. During the first session 8 of the 16 infants looked proportionately longer at the film accompanied by its asynchronous soundtrack; during the second session 13 of the 16 showed this effect.

It appears that infants were sensitive to the asynchrony in the expressive displays presented to them, but the altered temporal relations did not prevent them from detecting other invariant relations or common properties revealing which facial expression was accompanied by an affectively-appropriate vocal one. While the infants participating in the first two experiments could have been using synchrony relations, it is possible for infants to select the appropriate visual display without precise synchrony information. They were presumably guided by a common affordance (meaning) or some other factor that unified the optic and acoustic displays.

Experiment 4

The asynchrony experiment was aimed at exploring further the roles of temporal synchrony and common affective tone for the infants' discrimination of expressions. It demonstrated that precise temporal synchrony alone could

Not account for the infants' looking proportionately longer at a filmed facial expression when it was accompanied by its (asynchronous) soundtrack. But how essential is the common meaning of the optic and acoustic displays?

In the final experiment of this series, infants were presented films of angry and happy facial expressions accompanied by a single vocal expression. While the happy and neutral, and happy and sad pairs that were used before were visually quite distinct, the amount of facial movement in these expressions is comparable, although the tenor of each is, of course, quite different. For half of these seven-month-old infants, procedures were identical to those used in the earlier experiments. For the remaining half the films were presented side by side as before, but upside-down. By showing the faces upside-down all of the physical pattern information (other than orientation) was preserved, but perceiving the configurational aspects of the face required for the appreciation of facial expression became more difficult.

Those infants who saw the faces in a normal orientation looked longer at a film when it was in sound, compared to when it was silent, as in the preceding experiments (see Figure 2). These infants looked to the angry film 33% of the time when it was silent, 51% when it was in sound [.509 compared to .328, $t(15) = 3.62$, $p < .0025$]; they looked to the happy film 49% of the time when it was silent, and 67% when it was in sound [.673 compared to .491, $t(15) = 3.62$, $p < .0025$]. They also looked longer to the happy film (whether it was silent or sound-specified); that proportion averaged .582, $t(31) = 2.50$, $p < .018$. In contrast, those infants who were presented upside-down facial expressions along with a single soundtrack did not show any differences in looking times to the films.

Discussion

The conclusion seems fully warranted that infants detect information that is invariant across optic and acoustic displays of a single, affective expression. The five- and seven-month-olds in these experiments consistently increased their visual exploration of the expression that they heard as compared with the same expression without sound. Since several kinds of information are invariant over the optic and acoustic presentations of an affective expression, can we determine what information the infants are using? In addition, what do the results of these experiments tell us about the perception of the affordances or meaning of expressive behaviors?

Temporal synchrony relations are the most obvious type of invariants available in multimodally-presented expressive behaviors. The articulatory movements and vocal structure of an expression are concordant; that is, the sound of a laugh co-occurs with a wide open mouth, silence is usually accompanied by a closed mouth, and some syllables are clearly formed by pursing or opening the lips. The audible and visible behaviors also share a common, if irregular, tempo. However, the asynchrony experiment demonstrated that, while infants can detect the discrepancy in an asynchronous pairing of a filmed facial expression and its characteristic soundtrack, they detect other information that specifies a single, affective expression as well. Furthermore, in my final experiment in the inverted images condition temporal synchrony alone did not insure appropriate looking by infants. What other information for unity is available?

What one might refer to as "intensity relations" may exist between the optic and acoustic displays. That is, the animation visible in the happy

and angry facial expressions is paralleled by the wide range of pitch and loudness found in the accompanying vocal expressions. The relative stillness evident in the sad and neutral facial expressions is approximated by the "flat" tone of voice of the appropriate vocalizations. This explanation--that the infants detected intensity relations--does not easily account for the results of the normal orientation condition of Experiment 5, however. Infants in that experiment selected the facial expression that was accompanied by its soundtrack quite readily but the happy and angry facial expressions resembled one another closely as to amount of facial movement. Further, there is no evidence that vocal intensity and marked changes in facial expression are, in fact, highly correlated.

A more promising possibility is that the infants perceived a facial expression and its vocal counterpart as specifying the same meaning. By five months infants have had many opportunities to view the expressive behaviors of adults around them, and to learn what other behaviors these expressions afford. Perhaps the infants were able to detect the meaning that was invariant across the optic and acoustic presentations of happy, sad, neutral, and angry expressions. The results of Experiments 3 and 4 point directly to this as an explanation. Seven-month-old infants looked preferentially at the sound-specified happy and neutral films even when the accompanying vocal expression was out-of-synchrony. Synchrony obviously played a role in their looking patterns for they looked essentially randomly during the first session, but consider the perceptual task presented to them. Normally a speaker's face and voice are spatially congruent and are synchronized temporally, but neither was true in this experiment. The infants had to discard these typically dependable sources of information and rely

on something else in order to perceive the concordance between optic and acoustic displays. That the infants did not look at the sound-specified film during the first session indicates that asynchrony was detected and that time is necessary for determining which relations are relevant when those available are in conflict or unreliable. In the upside-down faces experiment precise temporal synchrony was present but did not appear to influence the infants' looking preferences when the filmed faces were inverted. All of the physical pattern information (other than orientation) was identical for the two groups of infants and the molecular psychophysical correspondences remained the same, yet the infants looked preferentially only when the faces were presented in an upright orientation.

To say that infants perceive the meaning common to two expressive displays can be interpreted in several ways. First, it is reasonable to suspect that infants of five and seven months could have learned to associate smiling faces with cheery voices or lowered brows with gruff or strident voices since these patterns typically co-occur. When these infants heard a gruff voice they may have simply searched for a face with knitted, lowered brows. All that would be necessary for appropriate looking is the experience of seeking and hearing the two together in the past. This interpretation is credible but cannot easily explain the results of the asynchrony experiment. If the infants simply were matching modality-specific representations why would the asynchronous pairings of the happy and neutral faces and voices disrupt preferential looking?

In view of these difficulties I prefer another account of the development of the perception of expressive behaviors. It is based on E. J. Gibson's

(1969) theory of perceptual development, and has as its fundamental mechanisms increasing differentiation and abstraction of invariant relations in multimodally-presented events. The results of these four experiments are consistent with this view and support the notion of the detection of affordances as espoused by James J. Gibson (1979). According to this view, infants become progressively more skilled at detecting invariants that specify significant properties or events in the environment (E. J. Gibson, 1969; J. J. Gibson, 1979; Spelke, 1979; in prep.). In the experiments reported here the infants detected a number of auditory-visual relationships, thereby perceiving the bimodally-presented expressions as unitary events. The infants were sensitive to temporally invariant information in the filmed expressions: asynchronous presentation affected their looking time to the sound-specified and silent facial expressions. But the infants were also sensitive to invariant information specifying the affordance of the facial and vocal expressions. These invariants seemed primary to their perception of unified expressions, with the temporal invariants being a unifying condition but not specific to meaning. Exactly how to characterize these invariants is not yet clear. We do know that the information available optically appears to require an upright presentation of facial expression. Further experiments are underway to clarify the nature of such information.

Finally, the data reported here indicate that infants are sensitive to invariant information specifying the affordances of affective expressions, but they do not directly address questions about the infants' expectations. Would an infants expect the genuine smiling, laughing actress to pick her up, or that same frowning, curt actress to reject her? To answer this question,

infants' affective reactions to these films were videotaped and have been coded as part of a program of research studying the development of perception of interpersonal affordances. The infants do behave in detectably different ways to happy and sad expressions, lending more credence to the idea that infants perceive multimodally presented expressions as meaningful behavior. The early sensitivity to intermodal relations in complex, expressive events surely enables the maturing child to progress rapidly in his social interactions and emotional development.

References

Allen, T. W., Walker, K., Symonds, L. and Marcell, M. Intrasensory and intersensory perception of temporal sequences during infancy. Developmental Psychology, 1977, 13, 225-229.

Bahrick. L. E., Walker, A.S. and Neisser, U. Selective looking by infants. Cognitive Psychology, in press.

Dodd, B. Lipreading in infants: attention presented in-and out-of-synchrony. Cognitive Psychology, 1979, 11, 478-484.

Gibson, E. J. Principles of perceptual learning and development. New York: Appleton-Century-Crofts, 1969.

Gibson J. J. An ecological approach to visual perception. Boston: Houghton-Mifflin, 1979.

Spelke, E. Infants' intermodal perception of events. Cognitive Psychology, 1976, 8, 553-560.

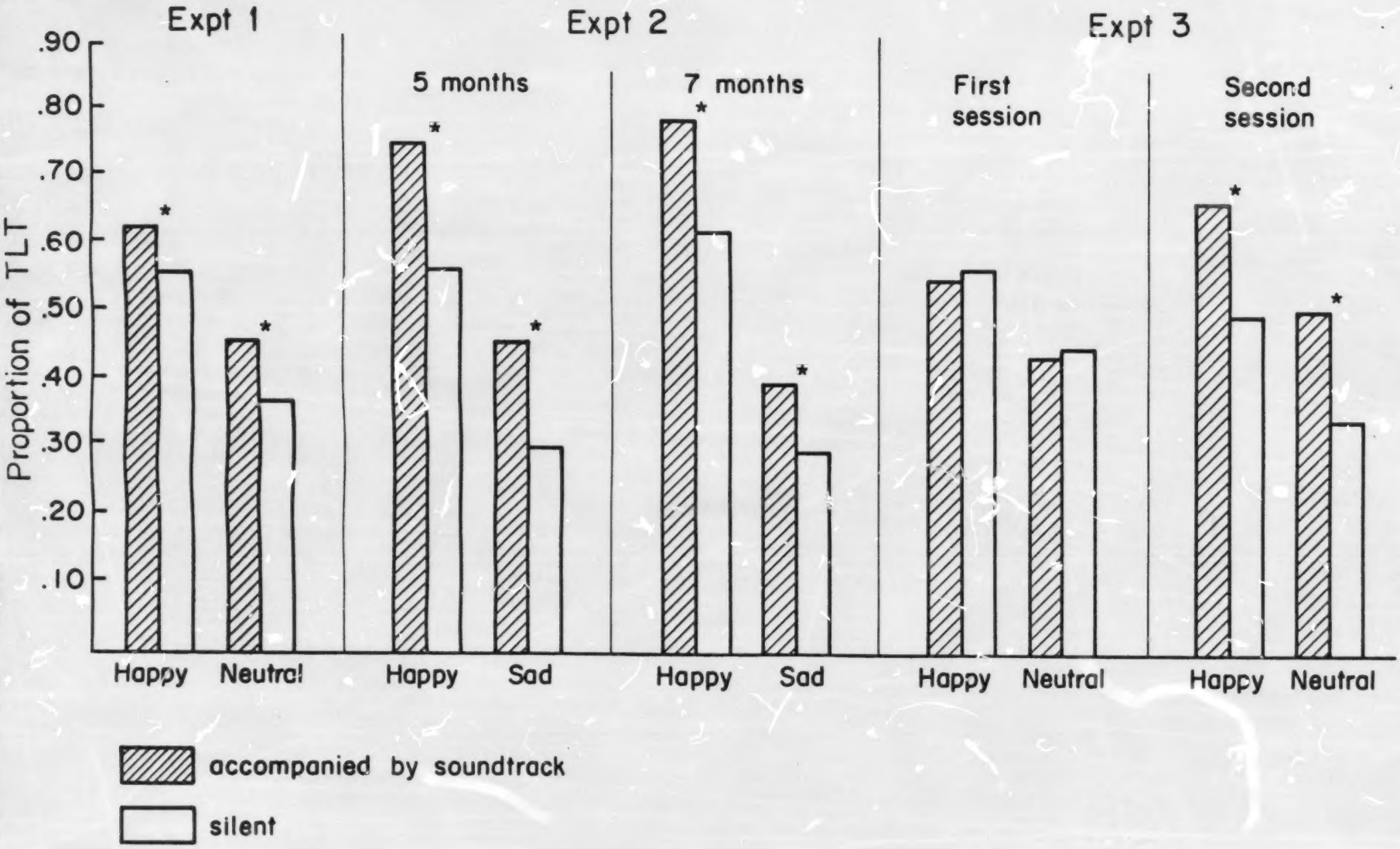
Spelke, E. S. Perceiving bimodally specified events in infants. Developmental Psychology, 1969, 15, 626-636.

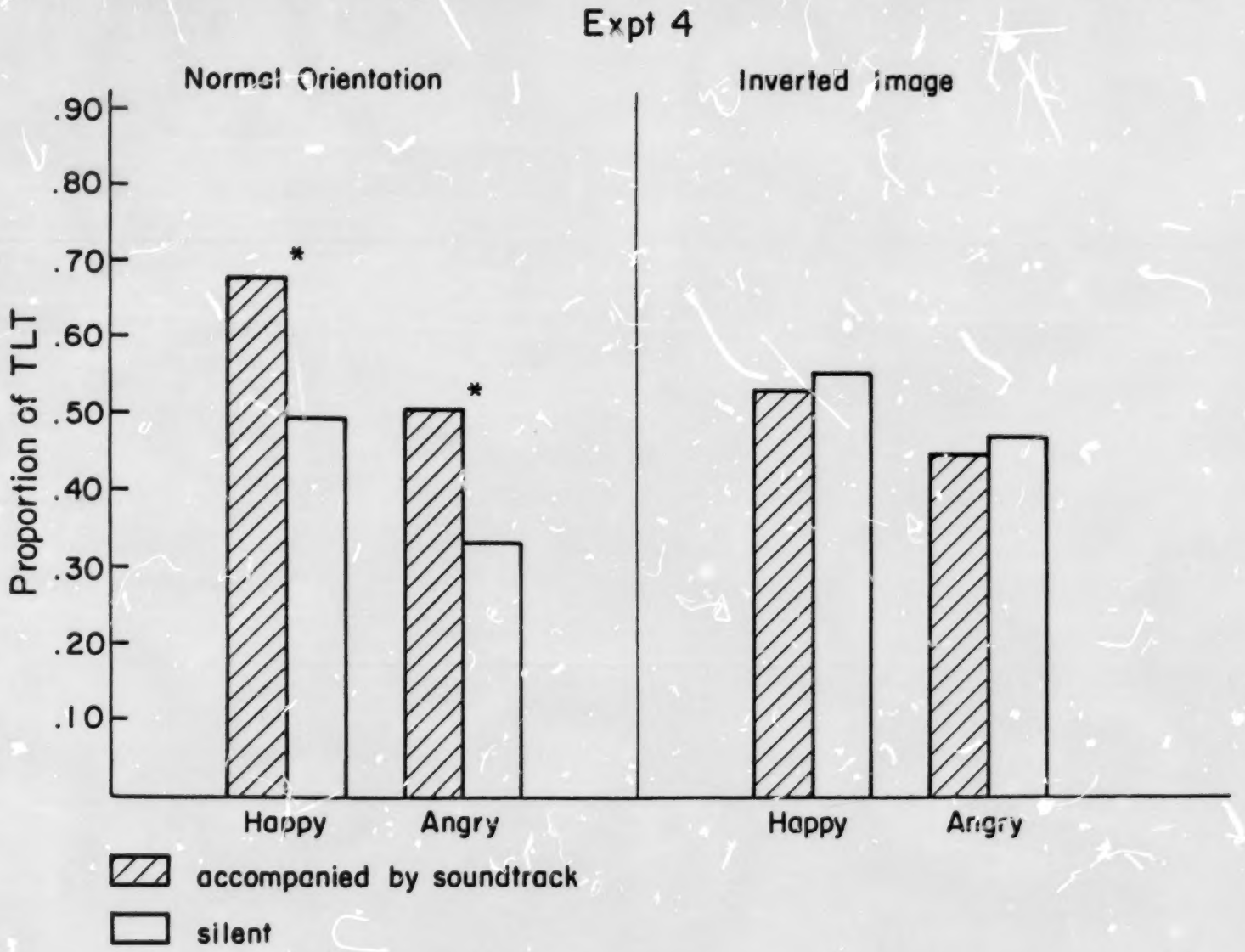
Spelke, E. S. The development of intermodal perception. In L. B. Cohen and P. Salapatek (Eds.), Hardbook of infant perception, in preparation.

Spelke, E. S. and Cortelyou, A. perceptual aspects of social knowing: looking and listening in infancy. In M.E. Lambard L. R. Sherrod (Eds.), Infant social cognition. New Jersey: Erlbaum and Associates, 1980.

Wilcox, B. and Clayton, F. Infant visual fixation on motion pictures of the human face. Journal of Experimental Child Psychology, 1968, 6, 22-32.

Young-Browne, G. Rosenfeld, H.M. and Horowitz, F. D. Infant discrimination of facial expressions. Child Development, 1966, 48, 555-562.


Expt 1
Expt 2
Expt 3
5 months
7 months
First session
Second session
Proportion of TLT
.90
.80
.70
.60
.50
.40
.30
.20
.10
Happy
Neutral
Happy
Sad
Happy
Sad
Happy
Neutral
Happy
Neutral
accompanied by soundtrack
silent


Expt 4
Normal Orientation
Inverted Image
Proportion of TLT
.90
.80
.70
.60
.50
.40
.30
.20
.10
*
*
Happy
Angry
Happy
Angry
accompanied by soundtrack
silent